

EFC3 - MLP e SVM

Rafael Gonçalves - RA: 186062

12 de Junho de 2019

Parte I - Revisitando o algoritmo de retropropagação de erro

Parte II - Atividade computacional

a) Solução fechada com MMQ

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \quad \Phi = \Phi(\mathbf{X}) = \begin{bmatrix} 1 & \mathbf{x}_1 \\ 1 & \mathbf{x}_2 \\ & \vdots \\ 1 & \mathbf{x}_N \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Modelo:

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}(\mathbf{X}) = \Phi \mathbf{w} \quad (1)$$

Solução ótima com o método de mínimos múltiplos quadrados (MMQ):

$$\mathbf{w} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{y} \quad (2)$$

Raiz quadrada do erro quadrático médio (RMSE) para o modelo treinado com o conjunto de treino:

$$RMSE_{train} = 15.3702 \quad RMSE_{test} = 14.2495$$

Resultado no conjunto de teste:



Figura 1: Saída prevista pelo modelo (azul) em comparação com valores reais dos últimos 5 anos (60 meses).

b) Seleção de atributos usando wrapper (backward elimination), validação cruzada (k-fold) e regularização L2

Solução por MMQ com regularização ridge (L2):

$$\mathbf{w} = (\mathbf{\Phi}^T \mathbf{\Phi} + \lambda \mathbf{I}')^{-1} \mathbf{\Phi}^T \mathbf{y} \quad (3)$$

Modelo escolhido pelo wrapper utilizando k-fold com 5 pastas:

N de atributos removidos: 9 $\lambda : 10000$ $RMSE_{CV} : 15.4201$ $RMSE_{test} : 14.4200$

Atributos escolhidos: 0 (bias) , 1, 2, 3, 4, 5, 6, 8, 9, 11, 18, 20

c) Seleção de atributos usando filtro (correlação de Pearson)

Modelo escolhido pelo filtro (11 atributos com maior correlação):

Atributos selecionados: 0 (bias) , 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

$\lambda : 10000$ $RMSE_{CV} : 15.7380$ $RMSE_{test} : 13.8367$

Ambas as estratégias de seleção de atributos em conjunto com a regularização L2 mostraram melhoria em relação ao uso de todos os 20 atributos sem regula-

Tabela 1: RMSE mínimo e respectivo lambda para cada número de atributos retirados.

N atributos retirados	min RMSE	lambda
0	15.4585	30000
1	15.4441	30000
2	15.4355	30000
3	15.4311	30000
4	15.4280	30000
5	15.4255	30000
6	15.4238	10000
7	15.4226	10000
8	15.4212	10000
9	15.4201	10000
10	15.4202	10000
11	15.4216	10000
12	15.4268	10000
13	15.4392	6000
14	15.4893	6000
15	15.5565	1000
16	15.7354	0
17	16.0103	0
18	16.0801	0
19	16.8718	0



Figura 2: Menor RMSE para cada passo do wrapper (atributo retirado).



Figura 3: Saída prevista pelo modelo (azul) em comparação com valores reais dos últimos 5 anos (60 meses) com seleção de atributos usando wrapper.

rização. Podemos observar que embora o erro de validação tenha sido menor na abordagem com seleção de atributos usando wrapper, o erro no conjunto de testes foi menor na abordagem de filtro. Isso pode ser consequência de o modelo utili-

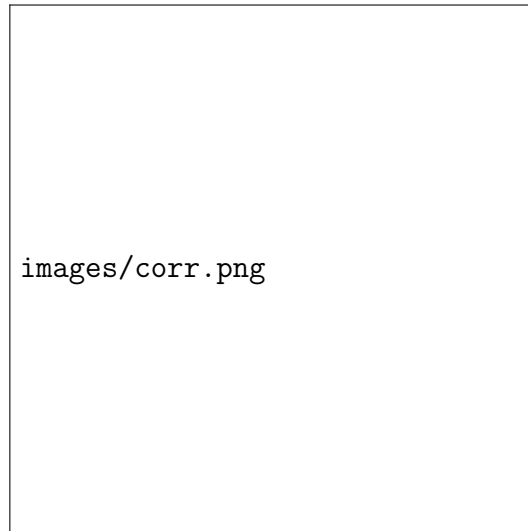


Figura 4: Matriz de correlação, onde uma cor mais avermelhada significa uma correlação maior (dados foram alinhados na matriz de forma que a primeira linha e coluna correspondem ao rótulo e as próximas linhas e colunas correspondem aos atributos - mês anterior até 20 meses atrás).



Figura 5: Saída prevista pelo modelo (azul) em comparação com valores reais dos últimos 5 anos (60 meses) com seleção de atributos usando filtro.

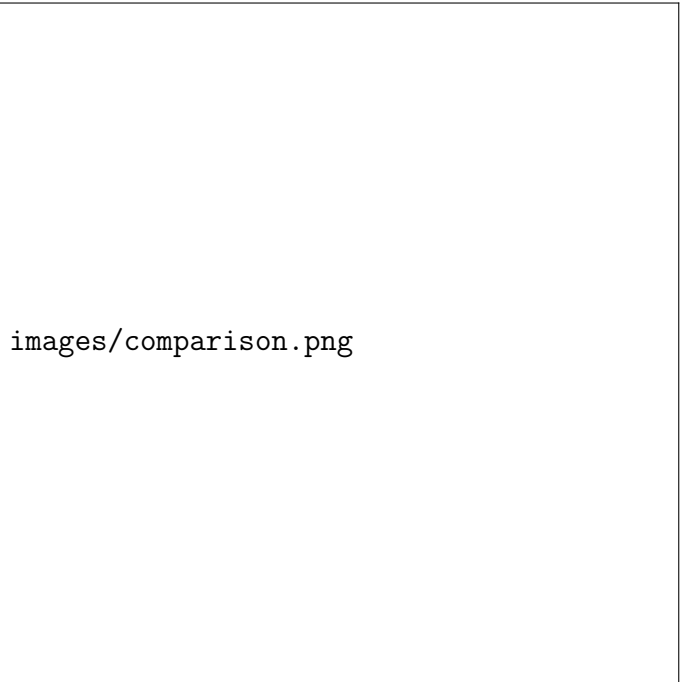


Figura 6: Saída prevista por cada modelo: azul - modelo inicial com 20 atributos, $RMSE = 14.2495$; amarelo - modelo criado usando wrapper, $RMSE = 14.1138$; verde - modelo criado usando filtro, $RMSE = 13.8367$; em comparação com valores reais (vermelho) dos últimos 5 anos (60 meses).

zado ser linear e portanto se beneficiar diretamente por atributos que tenha uma correlação com os rótulos.

Também vale notar que o modelo gerado com seleção de atributos pelo wrapper leva em consideração o erro da validação para escolher quais atributos usar, enquanto que o modelo gerado pelo filtro não leva em conta o erro da validação para seleção de atributos, apenas a correlação entre cada atributo e o rótulo e, sendo assim, faz sentido que o modelo do wrapper apresente um erro menor de validação em comparação ao modelo gerado usando filtro. Em todo o caso o resultado nos mostra que o modelo que obteve um melhor desempenho em dados novos foi o modelo gerado com o filtro e é o modelo que deve ser usado em uma aplicação - provavelmente terá melhor desempenho.